

Kanonická korelace

Menu:	QCExpert	Kanonické korelace
-------	----------	--------------------

Modul *Kanonická korelace* hledá obecný lineární vztah mezi dvěma vícerozměrnými proměnnými X a Y s obecně různými dimenzemi m_1, m_2 . Jedná se o obecnější metodu, než je párová a vícenásobná korelace v modulu *Korelace* založenou na projekci do hlavních komponent a nalezení takové lineární kombinace prvků první a druhé proměnné, která má maximální korelační koeficient. Metoda poskytuje statistický test významnosti vypočítané kanonické korelace, kanonické korelační koeficienty, kanonické proměnné a další výsledky. Cílem metody je podrobně identifikovat nejsilnější statistické vztahy mezi skupinami proměnných a pomoci uživateli při hledání skutečných příčinných vztahů. Výsledkem výpočtu jsou nové páry (dvojice) jednorozměrných proměnných A_i, B_i . Těchto párů je celkem $m = \min(m_1, m_2)$. Nejdůležitější je obvykle první pár

$$A_1 = a_{1,1}x_1 + a_{2,1}x_2 + \dots + a_{m_1,1}x_{m_1}$$
$$B_1 = b_{1,1}y_1 + b_{2,1}y_2 + \dots + b_{m_2,1}y_{m_2}$$

který je sestavený tak, aby mezi těmito kanonickými proměnnými A_1, B_1 byl maximální možný párový korelační koeficient. Každá další kanonické proměnná A , resp. B je vždy ortogonální ke všem předchozím kanonickým proměnným A , resp. B : $A_i^T \cdot A_j = 0$; $B_i^T \cdot B_j = 0$ pro $i \neq j$. Potvrdí-li pak test statistickou významnost korelace, lze říci, že mezi skupinami 1 a 2 byla statisticky prokázána závislost na uživatelem zadané hladině významnosti α (obvykle $\alpha=0.05$). Znaménka kanonických korelačních koeficientů nemají význam.

Data a parametry

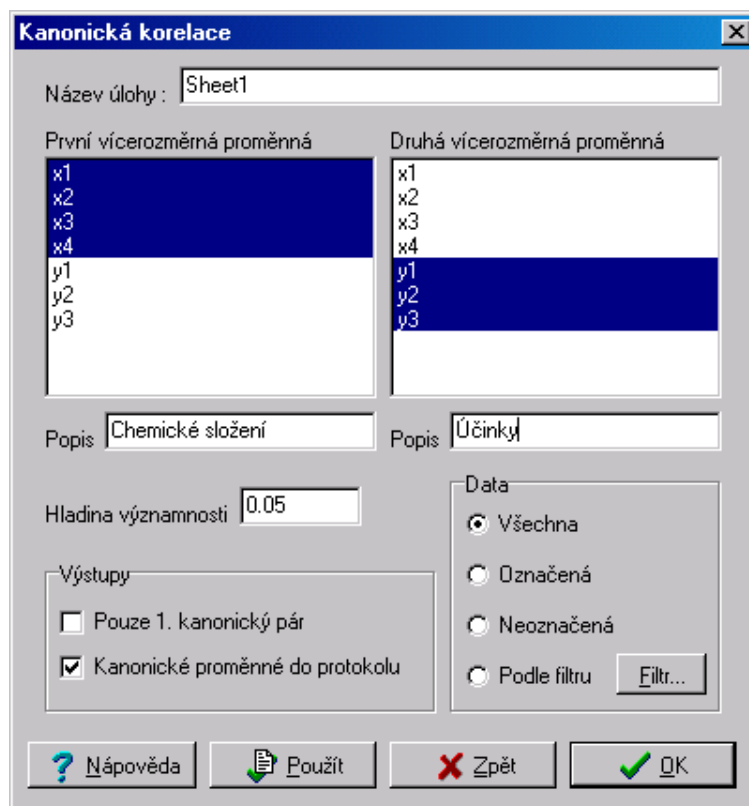
Analyzují se dva vícerozměrné výběry na základě dat uspořádaných do dvou skupin sloupců, jak je naznačeno v následující tabulce Tab.1. Dvě uživatelem zvolené skupiny sloupců obvykle charakterizují dvě skupiny veličin, mezi nimiž očekáváme určitou souvislost. V Tab.1 představují tedy první skupinu sloupce x_1, x_2, x_3, x_4 a druhou skupinu sloupce y_1, y_2, y_3 . Počty proměnných m_1, m_2 ve skupinách mohou být rozdílné a musí být větší než 1. Hodnoty v řádku musí odpovídat vždy témuž vzorku, situaci, pacientovi, apod. V každém řádku musí být všechny hodnoty, řádek s chybějící hodnotou bude ignorován. Typickými skupinami parametrů mohou být například 1. skupina: chemické složení, 2. skupina: fyzikální parametry, nebo účinek; 1. skupina: parametry vstupní suroviny, 2. proměnná: parametry produktu; 1. skupina: výsledky psychologických testů, 2. skupina: známky ve škole; atd.

Tab. 1 Příklad dat pro kanonickou korelaci

Vzorek č.	x1	x2	x3	x4	y1	y2	y3
33	8.08	2.89	500	21	6.5	28	5.24
34	8.29	4.43	600	22	6.1	32	6.51
35	8.81	3.92	600	19	5.7	33	7.91
36	8.53	3.75	700	17	5.1	38	8.15
37	9.04	3.77	600	12	3.4	32	7.02
39	7.44	2.5	500	15.5	3.8	27	6.255
44	8.83	3.46	500	14.5	4.1	28	6.555
45	7.82	3.2	600	22	4.9	33	5.94
48	8.43	3.31	500	14.5	4.1	28	6.125
15	8.02	2.9	500	21.5	5	27	5.125
16	8.91	3.08	500	21	5.2	29	6.54
17	8.95	3.14	600	17.5	4.8	33	7.855
18	8.88	3.69	600	17	4.2	32	7.64
19	7.28	4.08	600	21	5.2	32	7.51

Po otevření dialogového okna *Kanonická korelace* se vyberou sloupce první a druhé proměnné. Vybrané sloupce se nesmí překrývat! Sloupec vybraný v jedné skupině se nesmí objevit ve druhé skupině. Můžeme zadat popis první a druhé skupiny a hladinu významnosti (obvyklá hodnota hladiny

významnosti je 0.05 (tedy 5%). Dále lze specifikovat obsah výstupního protokolu. Zaškrtneme-li políčko *Pouze první kanonický pár*, bude ve výsledku uveden pouze první pár kanonických proměnných A_1, B_1 . Nezaškrtneme-li políčko *Kanonické proměnné do protokolu*, nebudou v protokolu hodnoty kanonických proměnných uvedeny nezávisle na zaškrtnutí políčka *Pouze první kanonický pár*, což je výhodné při větším počtu řádků, kdy by byl protokol dlouhý a nepřehledný.



Obrázek 1 Dialogové okno modulu **Kanonická korelace**

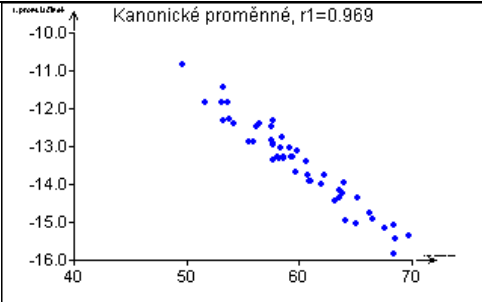
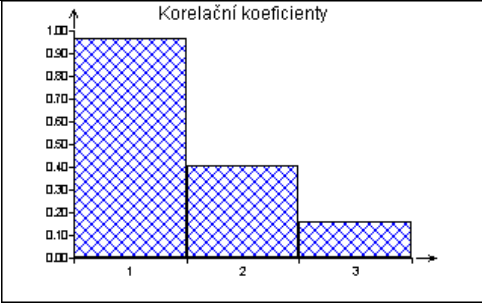
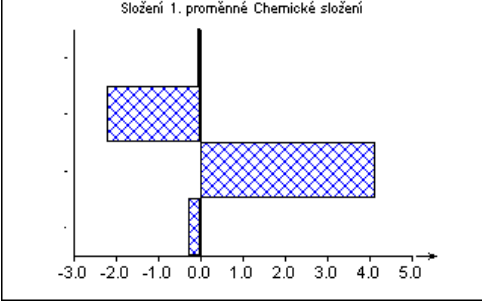
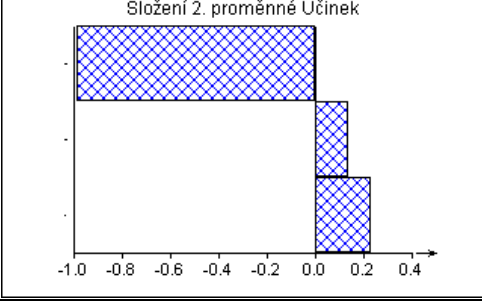
Ve skupině Data můžeme zvolit podskupinu dat podle označení v datové tabulce, případně definovat filtr. Po stisku tlačítka *Použít*, nebo *OK* se provede výpočet.

Protokol

Název úlohy	Název úlohy z dialogového panelu.
Data	Zvolená podmnožina dat
Základní charakteristiky	
i . vícerozměrná proměnná	Pořadí vícerozměrné proměnné ($i = 1, 2$)
Průměr	Aritmetický průměr jednotlivých sloupců i -té proměnné.
Směr. odchylka	Směrodatná odchylka jednotlivých sloupců i -té proměnné.
Kanonické korelační koeficienty	
Korelace i	Hodnota i -tého kanonického korelačního koeficientu, $i = 1, \min(m_1, m_2)$
Významnost korelace	Celková statistická významnost kanonické korelace
Lambda	Testová statistika s chí-kvadrát rozdělením

Chi kvadrát p-hodota Závěr	Kritický kvantil chí-kvadrát rozdělení pro danou hladinu významnosti p-hodnota testu Slovní závěr testu kanonického korelačního koeficientu (VÝZNAMNÝ, nebo NEVÝZNAMNÝ)
Složení kanonických proměnných	Koeficienty a_{ij} , b_{ij} pro výpočet kanonických proměnných
1. kanonické proměnná	Koeficienty a_{ij} , pro 1. kanonickou proměnnou
1. kanonické proměnná	Koeficienty b_{ij} , pro 2. kanonickou proměnnou
Hodnoty kanonických proměnných	Vypočítané hodnoty kanonických proměnných A_i , B_i . Tento odstavec bude chybět, nebylo-li zaškrtnuto políčko <i>Kanonické proměnné do protokolu</i>

Grafy

 <p>Kanonické proměnné, $r=0.969$</p>	Grafické reprezentace všech dvojic kanonických proměnných. Kritérium statistické významnosti je daleko přísnější, než u párové korelace, takže ani silně „vypadající“ závislost nemusí být statisticky významná. Znaménko korelace nemá význam.
 <p>Korelační koeficienty</p>	Sloupcový graf absolutních hodnot jednotlivých kanonických korelačních koeficientů.
 <p>Složení 1. proměnné Chemické složení</p>	Grafické vyjádření složení 1. kanonické proměnné, tedy koeficientů $a_{1,1}$, $a_{2,1}$, ..., $a_{m,1}$
 <p>Složení 2. proměnné Účinek</p>	Grafické vyjádření složení 2. kanonické proměnné, tedy koeficientů $b_{1,1}$, $b_{2,1}$, ..., $b_{m,1}$